Introduction to Probabilistic Models

Seungwoo Schin

Coding the Mathematics Course, Fastcampus

Contents

I. Mutivariate Random Variables

I. Probability Distribution for Discrete Multivariate Random Variable

I.1 Definition

이산확률변수 X_i에 대해서, 다음과 같은 함수를 다변수 이산확률분포함수라고 한다.

$$p(x_1, x_2, ..., x_n) = P(X_i = x_i, i = 1, 2..., n) \tag{1}$$

이 때 p는 다음의 두 조건을 만족해야 한다.

- $\forall x_i \in domain(X_i), p(x_1, ..., x_n) > 0$
- $\sum_{\forall x_i} p(x_1, ..., x_n) = 1$

I.2 Probability Functions

Joint Cummulative Function 위와 같이 이산확률변수 X_i에 대해서, 다음과 같은 함수를 Joint Cummulative Function이라고 한다.

$$F(x_1, x_2, ..., x_n) = P(X_i \le x_i, i = 1, 2, ..., n) = \sum_{X_i \le x_i} p(x_1, x_2, ..., x_n) \tag{2}$$

Marginal Distributions 이산확률변수 X_i에 대해서, 다음과 같은 함수를 X_1의 marginal distribution 이라고 한다.

$$g(x_1) = \sum_{x_i, i \neq 1} p(x_1, x_2, ..., x_n) \tag{3}$$

II. Probability Distribution for Continuous Multivariate Random Variable

II.1 Definition

연속확률변수 X_i에 대해서, 다음과 같은 함수를 다변수 연속확률분포함수라고 한다.

$$p(x_1, x_2, ..., x_n) = P(X_i = x_i, i = 1, 2..., n) \tag{4}$$

이 때 p는 다음의 두 조건을 만족해야 한다.

- $\forall x_i \in domain(X_i), p(x_1, ..., x_n) > 0$
- $\sum_{\forall x_i} p(x_1, ..., x_n) = 1$

II.2 Probability Functions

Joint Cummulative Function 위와 같이 연속확률변수 X_i에 대해서, 다음과 같은 함수를 Joint Cummulative Function이라고 한다.

$$F(x_1, x_2, ..., x_n) = P(X_i \le x_i, i = 1, 2, ..., n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \tag{5}$$

Marginal Distributions 연속확률변수 X_i에 대해서, 다음과 같은 함수를 X_1의 marginal distribution 이라고 한다.

$$g(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p(x_1, x_2, ..., x_n) dx_2 dx_3 ... dx_n \tag{6}$$

III. Conditional Probabiltiy/Independence

III.1 Conditional Probability Functions

확률변수 A,B에 대해서 조건부확률$P(A|B)$는 다음과 같이 정의된다.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{7}$$

Discrete Case 이산확률분포의 경우, 다음 식이 성립한다.

$$P(A=a|B=b) = \frac{P(A=a,B=b)}{P(B=b)} \tag{8}$$

$$P(A|B=b) = \frac{P(A,B=b)}{P(B=b)} \tag{9}$$

IV. Multivariate Normal Distribution

Definition n차원 확률변수 $\vec{x}$는 다음의 joint pdf를 가질 때 Multivariate Normal Random Vector 이라고 한다.

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{2\pi^{-n}det(V)}}e^{\frac{1}{2}(\vec{x}-\vec{mu})^T V^{-1}(\vec{x}-\vec{mu})} \tag{10}$$

이 때,

- $\vec{mu} = \frac{\sum_i X_i}{n}$
- $V_{ij} = cov(X_i, X_j)$

로 정의된다. 이런 확률변수 $\vec{x}$의 경우, 몇 가지 흥미로운 성질을 보이는데

- Component 중 몇 개를 선형결합하더라도 Normal Distribution이다.
- 각각의 component가 모두 Normal Distribution이다.

단, Normal Distribution X,Y의 결합이 항상 MVN이 되지는 않는다. 즉 위의 역은 성립하지 않는다. 이는 Covariance Matrix가 invertible하지 않을 수 있기 때문이다.

II. Stochastic Process

I. Definitions & Terminology

Definition Stochastic Process는 집합 T의 임의의 원소에 대해서 대응되는 Random Variable의 모임 $\{X(\theta, t) | t \in T\}$를 뜻한다. 여기서 θ는 T와 상관없는 확률변수의 파라미터이다.

이 때 중요한 것은 $X(\theta, t)$는 엄연히 함수라는 점이다. 표기 시에는 사실상 random variable과 이에 의해서 생성된 값을 구분하여 쓰지 않지만, 이 둘의 차이를 이해하는 것은 중요하다. 프로그래밍으로 비유하자면, 다음 두 코드의 차이라고 볼 수 있을 것이다.

```python
# code 1
def func(a, b):
    print(a+b+1)

for i in range(10):
    func(i, 1)

# code 2
def generate_func(a,b):
    def func(c):
        print(a+b+c)
    return func

for i in range(10):
    generate_func(i, 1)(1)

# code 3
for i in range(10):
    generate_func(i, 1)(2)
```

위 두 code 1, code 2는 똑같은 결과를 만들어내지만, 내부 로직은 많이 다르다. code 1의 경우는 상수를 계산하여 출력하는 것이고, code 2의 경우는 함수에 값을 대입하여 만든 것으로 본질적으로 함수의 sequence에 값을 대입하여 출력한 것이기 때문이다. 이는 code 3를 보면 더 명확해지는데, code 3의 경우 code 2와 *본질적으로* 같은 sequence이나 대입하는 인자가 다르므로 다른 값을 출력하게 된다.[1] 이와 같이, stochastic process는 random variable, 즉 함수의 sequence이지 숫자의 sequence가 아님을 명심해야 한다.

Terminology Stochastic Process $\{X(\theta, t) | t \in T\}$에 대해서,

- Index Set은 집합 T를 말한다.
- State Space S는 $\cup_{t \in T} range(X(\theta, t))$를 말한다. 즉, Stochastic Process의 결과인 random variable이 가질 수 있는 값들 전체의 집합을 말한다.
- Increment는 어떤 T 안의 두 점 t_1, t_2에 대해서 $X(\theta, t_1) - X(\theta, t_2)$를 말한다.
- Ensemble은 $\{X(\theta, t) | \theta \in \mathbb{R}\}$ 을 말하며, Sample Function은 Ensemble의 원소를 말한다.
- Stationary Process는 임의의 T의 원소에 대해서, 모든 X(t)가 같은 확률 분포 함수를 가지는 process를 말한다.

[1] 이 때, code 2와 3은 각각 같은 stochastic process의 다른 sample function에 비유할 수 있다.

- Discrete Stochastic Process는 State Space T가 가산집합일 때를 말한다. 반대로, T가 비가산집합이면 Continuous Stochastic Process라고 한다.

용어들을 조금 더 쉽게 이해하기 위해서 Discrete Stochastic Process와 Continuous Stochastic Process의 예를 하나씩 들어 보고자 한다.

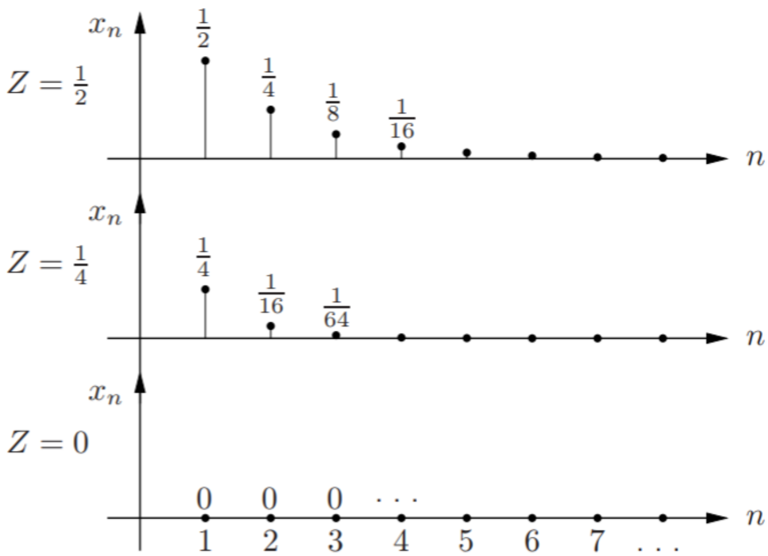

Figure 1: *Discrete Stochastic Process X_n*

Example II.1. $Z \sim U[0,1]$이라고 할 때, Discrete Stochastic Process X_n을 $X_n = Z^n$으로 정의하자. 이 때, X_n의 Sample Function은 Figure 1와 같다. 이 때, X_n은 Z와 n 둘의 함수임에 유의하라. Z가 상수일 때, X_n은 n만의 함수이며, 이 때 각각의 n에 대한 함수 X_n을 이 Stochastic Process의 Sample Function, 혹은 Sample Path라고 한다.

또한, 이 때 이 stochastic process의 pdf $f_{X_n}(x)$를 생각해 볼 수 있다. 즉,

$$f_{X_n}(x) = \frac{dF(x)}{dx} = \frac{dP(X_n \leq x)}{dx} \tag{11}$$

이다.

이 때, 아래와 같은 과정을 통해서 $f_{X_n}(x) = \frac{1}{nx^{(n-1)/n}}$임을 알 수 있다.

$$\begin{aligned}
P(X_n \leq x) &= P(Z^n \leq x) && (12)\\
&= P(Z \leq x^{\frac{1}{n}}) && (13)\\
&= x^{\frac{1}{n}} && (14)\\
\frac{dP(X_n \leq x)}{dx} &= \frac{d(x^{\frac{1}{n}})}{dx} && (15)\\
&= \frac{1}{n} x^{\frac{1}{n}-1} && (16)
\end{aligned}$$

이 때, 이 stochastic process의 pdf가 n에 대한 함수임을 볼 수 있다. 이런 경우, X_n은 stationary 하지 않다고 한다.

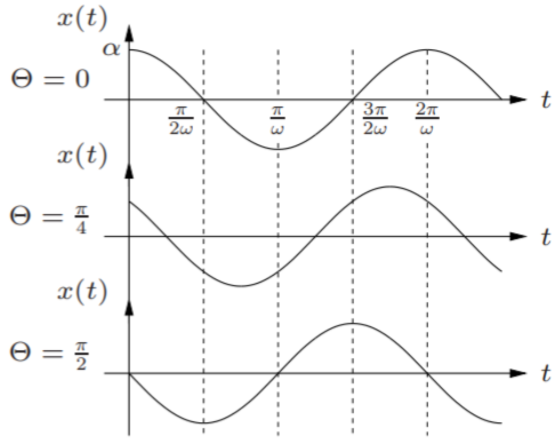

Figure 2: *Continuous Stochastic Process* $X(t)$

Example II.2. $\theta \sim U[0, 2\pi]$라고 할 때 Continuous Stochastic Process $X(t)$가 상수 α, ω 에 대해서 $X(t) = \alpha cos(\omega t + \theta)$ 로 주어졌다고 하자. 이 때 Sample Function은 Figure 2와 같다. 위 이산적인 경우와 비슷하게, X(t)는 t와 θ의 함수임에 유의하라. 이번에도 위와 비슷하게 pdf $f_{X(t)}(x)$를 구해 보자.

$$P(X(t) \leq x) \quad = \quad P(\alpha cos(\omega t + \theta) \leq x) \tag{17}$$
$$\quad = \quad P(\theta \leq arccos(\frac{x}{\alpha}) - \omega t) \tag{18}$$

여기서, $\frac{darccos(x)}{dx} = \frac{-1}{\sqrt{1-x^2}}$ 임을 이용하면 $f_{X(t)}(x) = \frac{1}{\alpha\pi\sqrt{1-(x/\alpha)^2}}$임을 알 수 있다. 이 때, pdf가 t에 dependent하지 않으므로, 이 stochastic process의 경우 t와 상관없이 pdf가 일정하게 유지된다. 이와 같은 stochastic process를 stationary process라고 한다.

위 예제들에서 알 수 있듯이, $X(\theta, t)$와 θ의 분포를 직접적으로 특정하는 것으로 stochastic process를 기술할 수 있다. 하지만 대부분의 경우는 θ의 분포가 직접적으로 특정되지 않는 경우가 많다. 따라서 앞으로는 대부분의 stochastic process를 $\forall t \in T$에 대해서 X_t를 각각 특정하는 것으로 기술할 것이다.

II. Classes of Stochastic Processes

본 단락에서는 유명한 stochastic process들의 성질들을 살펴보고자 한다. 이 성질들을 만족하는 process들은 보통 그 class의 stochastic process라고 한다.

II.1 Identical/Independent Processes

Definition 가장 기본적인 Stochastic Process는 $\forall t \in T, X_t$가 iid, 즉 독립이며 같은 확률변수인 경우이다. 이러한 process를 IID Stochastic Process라 한다. 예로는

- Bernoulli Process : $X_i \sim Bernoulli(p)$
- Discrete-time White Gaussian Noise(WGN) : $X_i \sim N(0, \sigma)$

등이 있다.

Random Walk with Constant Step Random Walk Process는 정확하게는 IID Process가 아니지만, 연관이 있기에 여기에 설명한다. Random Walk Process는 다음과 같이 정의된다.

$$X_0 = 0 \tag{19}$$

$$X_n = \sum_{i=1}^{n} Z_i \tag{20}$$

이 때, $Z_i = 2X_i - 1, X_i \sim Bernoulli(0.5)$이다. 즉, 0.5의 확률로 1, 0.5의 확률로 -1인 경우를 말한다. 직관적으로 볼 때, random walk process는 동전을 던지며 계단을 오르내리는 것과 비슷하다. 동전을 던져 앞면이 나오면 위로, 뒷면이 나오면 아래로 한 칸 움직인다고 생각하자. 이 때 n번 던진 후에 몇 칸이나 위로 올라갔는지가 X_n이 될 것이다. 이런 경우 sample path는 $|a_n - a_{n+1}| = 1$이 모든 n에 대해서 성립하는 임의의 수열이 될 것이다. 예컨대, (0,1,2,3,....) 이나, (0, -1, 0, 1, 2, ...) 등은 random walk의 sample path이다.

Random Walk의 pmf $f_{X_i}(k) = P(X_i = k)$를 구해보자. 우선, X_i가 가질 수 있는 값에 대해서 생각해보자. 만약 i가 짝수라면, X_i의 값 또한 짝수어야 할 것이다. 또한, 반대로 i가 홀수라면 X_i의 값 또한 홀수여야 할 것이다. 즉, $i + X_i$는 언제나 짝수여야 하며, 또한 X_i는 i와 -i의 사이에 있을 것이다. 즉,

$$f_{X_i}(k) = 0, \text{ for } |k| > i, i + k \equiv_2 1 \tag{21}$$

이제 i+k가 짝수인 경우에 대해서 생각해 보자. 만약 1이 나온 횟수를 a라고 하면, -1이 나온 횟수는 i-a이므로 k=a+(-1)(i-a)가 성립하여, $a = \frac{i+k}{2}$이다. 따라서 Binomial Distribution에서의 확률과 비슷하게

$$f_{X_i}(k) = \binom{i}{\frac{i+k}{2}} 2^{-n} \tag{22}$$

이 된다.

II.2 Counting Processes

Definition Counting Process는 다음을 만족하는 Stochastic Process들을 말한다.

- $N(t) \geq 0$
- N(t)는 정수이다.
- $s \leq t \rightarrow N(s) \leq N(t)$

이름에서 알 수 있듯이, 어떤 특정한 사건이 t까지 몇번 일어났는지를 세는 등의 문제에서 많이 이용된다.

Poisson Process Poisson Process는 여러 분야에서 많이 활용되는 Stochastic Process로, Counting Process의 일종으로 볼 수 있다[2]. 만약 t를 시간으로 본다면 Timal Poisson Process라 하며, 큐 이론 (Queueing Theory)등에서 객체의 도착 등을 모델링할 때 사용된다. 만약 t가 어떠한 공간 내에서의 좌표라면, 무선네트워크나 감지기에서 특정 입자를 검출할 확률 등으로 사용할 수 있다.

Poisson Process는 어떤 하나의 수학적 object로 그 성질이 많이 결정되는데, 이 object는 대개의 경우 상수이거나 적분 가능한 함수이다. 보통 이 object는 λ로 표기하며, 상수인 경우 homogeneous/stationary Poisson Process라 하고 그 process의 rate나 intensity를 나타내게 된다. 반대로 적분 가능한 함수일 경우 inhomongeneous Poisson Process라고 하며, 여기서는 자세히 다루지 않는다.

[2] 이는 Poisson Process의 유일한 해석이 아니다. 하지만 본 수업에서는 이 해석을 위주로 살펴볼 것이다.

Poisson Distribution and Exponential Distribution 이름에서 알 수 있듯이, Poisson Process는 Poisson Distribution과 밀접한 관계[3]를 가지며, Exponential Distribution과도 관련이 있다. 따라서 Poisson Distribution과 Exponential Distribution을 다시 살펴보자. 두 확률분포는 다음과 같은 pdf를 가진다.

$$P_{poisson}(N=n) = \frac{\lambda^n}{n!}e^{-\lambda} \tag{23}$$
$$P_{exp}(N=n) = \lambda e^{-\lambda x} \tag{24}$$

	μ	σ^2	cdf
Poisson	λ	λ	-[4]
Exponential	λ^{-1}	λ^{-2}	$1-e^{-\lambda x}$

Table 1: *Statistics about Poission and Exponential Distributions*

여기서 두 확률분포의 평균과 분산, 그리고 cdf는 위 표에 나와있다.

Memoryless Property Memoryless Property라 함은 확률변수 T가 아래의 식을 만족함을 말한다.

$$P(T > s+t | T > s) = P(T > t) \tag{25}$$

만약 T를 어떤 사건(예를 들어서, 어떤 기계의 고장 등)이 일어날 때까지 걸리는 시간이라고 해 보자. 이 때 위 Memoryless Property에서의 각 항은 다음과 같이 해석될 수 있다.

- P(T>s) : 어떤 사건이 일어날 때까지 걸리는 시간이 s 이상일 확률. 즉, s동안은 어떤 사건이 일어나지 않을 확률.
- P(T>s+t | T>t) : t동안 어떤 사건이 일어나지 않을 때, s+t동안 그 사건이 일어나지 않을 확률.

예를 들어서, 전구가 10년동안 고장나지 않았다고 하자. 이 때, 전구가 고장날 때까지의 시간이 Memoryless Property를 따르는 확률변수라면 20년 후에 고장날 확률은 똑같은 새 전구가 10년 후에 고장날 확률과 같다.

Memory Property를 가지는 확률변수로는 Exponential Distribution이 있다. 이는 Exponential Distribution의 cdf를 이용하여 다음과 같이 쉽게 보일 수 있다.

$$\begin{aligned}
P(T > s+t | T > s) &= \frac{P(T > s+t \cap T > s)}{P(T > s)} &(26)\\
&= \frac{P(T > s+t)}{P(T > s)} &(27)\\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} &(28)\\
&= e^{-\lambda t} = P(T > t) &(29)
\end{aligned}$$

즉, Exponential Distribution은 Memoryless Property를 가진다.

역으로, 어떤 분포가 Memoryless Property를 가지고 정의역이 양수라면 그 분포는 Exponential Distribution이다. 이는 다음과 같이 보일 수 있다. 먼저, Memoryless Property가 다음의 두 성질을 imply함을 보이자. 이 때 F는 $F(x) = P(T > x)$로 정의된다.

[3]다만, Poission Process는 수학자 Poisson과는 무관하다.

- $F(a+b) = F(a)F(b)$: Memoryless Property의 정의에서 자명하다.
- F는 단조감소함수 : F의 정의에 의해서 자명하다.

이 두 성질을 이용하면, 아래 식이 성립함을 보일 수 있다.

$$F(x) = F(1)^x \tag{30}$$

위 식은 x가 정수라면 자명하고, x가 정수가 아니라고 하더라도 모든 실수 x에 대해서 성립[5]한다. 따라서 F(x)는 지수함수가 되며, 이를 미분한 함수가 pdf이므로 Exponential Distribution임을 보일 수 있다.

Poisson Process as Counting Process 이제 본격적으로 Poisson Process를 살펴보자. 어떤 Counting Process N(t)는 다음의 3가지 성질을 만족할 때 Poisson Process이다.

- N(0) = 0
- 임의의 index set T 안의 증가수열 n_i에 대해서, $X_{n_{i+1}} - X_{n_i}$가 독립임. (Independent Increment)
- index set T에 대해서 $\forall k \in T, N(t+k) - N(k) \sim Poisson(\lambda t)$

마지막 성질에서 k를 0으로 잡아보면, N(t)는 다음과 같은 성질을 가진다.

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{31}$$

또한, N(t)가 1만큼 증가할 때 t의 간격은 Exponential Distribution이 된다. 이를 inter arrival time 이라고 하며, 그 분포가 Exponential Distribution임을 아래와 같이 보일 수 있다. 먼저, N(t)=0이 유지될 시간에 대한 분포를 생각해 보자. Poisson Distribution의 정의에 따라 아래 식이 성립한다.

$$P(N(t) = 0) = e^{-\lambda t} \tag{32}$$

맨 처음 이벤트가 일어나는 시간을 T_1이라고 하면, $P(N(t)) = P(T_1 > t)$라고 볼 수 있으므로 T_1은 Exponential Distribution임을 알 수 있다. 이제, $s < t$인 상황에서 다음을 생각해 보자.

$$\begin{aligned}
P(N(s) = n, N(t) = m) &= P(N(s) = n, N(t) - N(s) = m - n) && (33)\\
&= P(N(s) = n)P(X(t-s) = m - n) && (34)\\
&= \frac{(\lambda s)^n}{n!} e^{-\lambda s} \frac{(\lambda(t-s))^{m-n}}{(m-n)!} e^{-\lambda(t-s)} && (35)\\
&= \frac{(\lambda t)^m}{m!} e^{-\lambda t} \binom{m}{n} \frac{s^n}{t^n} (1 - \frac{s}{t})^{m-n} && (36)
\end{aligned}$$

첫 번째 줄에서 두 번째 줄로 넘어갈 때는 Poisson Process의 independent increment를 사용하였다. 위 식을 이용하여 m=n인 경우를 보면,

$$P(N(s) = N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda(s+t)} \frac{s^n}{t^n} = \frac{(\lambda s)^n}{n!} e^{-\lambda(s+t)} \tag{37}$$

이 된다. 따라서

$$P(N(t) = n | N(s) = n) = \frac{P(N(s) = N(t) = n)}{P(N(s) = n)} = e^{-\lambda(t-s)} \tag{38}$$

가 된다.

[5]여기서는 증명은 생략하나, 증명 전략은 다음과 같다. 우선, 유리수인 경우에 $x = \frac{m}{n}$임을 이용하여 보이고, 실수인 경우에는 그 특정 실수로 좌, 우에서 수렴하는 유리수의 수열을 잡아서 샌드위치 정리를 이용하여 보일 수 있다.

II.3 Markov Processes

위에서 살펴본 IID 모델의 경우, process 내에서 각 X_i들이 서로 independent하였다. 이번에는 과거 X_i 들이 그 후의 stochastic process의 결과에 영향을 미치는 경우를 살펴보고자 한다. 즉,

$$P(X_i = k|X_j = a_j, j < i) \neq P(X_i = k) \tag{39}$$

일 때를 생각해보자.

Markovian Property 먼저, 문제를 간단하게 만들기 위해서 X_i가 이산적인 값을 가지면서 동시에 index set이 가산집합인 경우를 생각하자. 이 때,

$$P(X_n = i|X_{n-1} = j) = P_{ij} \tag{40}$$

로 정해질 때 X_i는 Markovian Property를 가진다고 한다. 이 때 P는 가질 수 있는 모든 state가 n개라면, n by n 행렬로 생각할 수 있다.

Discrete-Time Markov Process 위와 같이, X_i의 값이 이산적이면서 동시에 index set이 가산집합인 discrete stochastic process를 Markov Process라고 한다. 이 때 위에서 나온 행렬 P를 transition matrix 라 하며, state간의 transition 확률을 나타내는 행렬로 볼 수 있다.

Example II.3. Random Walk의 경우를 다시 살펴보자. 이 때 $P_{i,i+1} = 0.5$ 로, 다음과 같은 식의 무한 행렬로 생각할 수 있다.

$$\begin{bmatrix} 0 & 0.5 & 0 & 0 & ... \\ 0.5 & 0 & 0.5 & 0 & ... \\ 0 & 0.5 & 0 & 0.5 & ... \\ 0 & 0 & 0.5 & 0 & ... \end{bmatrix} \tag{41}$$

이와 같이 행렬로 나타낼 때의 장점은, transition이 여러 번 일어날 때 쉽게 확률을 계산할 수 있다는 점이다. n번 transtion을 다음과 같이 써 보자.

$$P_{ij}^n = P(X_{n+k} = j|X_k = i) \tag{42}$$

이 때, 다음 식이 성립한다.

$$P^(n+m)_{ij} = \sum_k P_{ik}^n P_{kj}^m \tag{43}$$

이를 Chapman-Kolmogorov Equation이라고 한다. 이 식이 성립하는 것은 다음과 같이 보일 수 있다.

$$\begin{aligned} P^(n+m)_{ij} &= P(X_{n+m} = j|X_0 = i) && (44) \\ &= \sum_k P(X_{n+m} = j, X_n = k|X_0 = i) && (45) \\ &= \sum_k P(X_{n+m} = j|X_n = k, X_0 = i)P(X_n = k|X_0 = i) && (46) \\ &= \sum_k P_{ik}^n P_{kj}^m && (47) \end{aligned}$$

만약 state의 갯수가 유한하다면, 이를 행렬 P의 곱으로 계산할 수 있다. 예컨대, 2번 transition의 경우를 보면

$$P_{ij}^2 = \sum k P_{ik} P_{kj} \tag{48}$$

인데, 이는 행렬 P의 제곱의 정의와 정확하게 일치하기 때문이다.

Classification of States

- Accessible State

  어떤 state i에서 state j로 언젠가 갈 확률이 있을 때, state j 가 접근 가능하다고 한다. 이 때 다음이 성립한다.

  $$\exists n P_{ij}^{n} > 0 \tag{49}$$

  이는 다음을 통해서 보일 수 있다.

  $$\begin{aligned} P(everenterj|starti) &= P(\cup_n X_n = j|X_0 = i) && (50)\\ &\leq \sum P(X_n = j|X_0 = i) && (51)\\ &= \sum P(X_n = j|X_0 = i) = 0 && (52) \end{aligned}$$

  이기 때문이다.

- Recurrent/Transient State

  어떤 state i에 대해서, 다시 state i로 올 확률을 p_i라 하면 $p_i = 1$이면 recurrent, $p_i < 1$이면 transient하다고 한다. 즉, transient한 state라는 뜻은 절대로 다시 원래 state로 돌아오지 않을 확률이 있다는 뜻이다. 직관적으로 생각할 때, transient 한 state는 유한번만 방문하게 된다고 생각할 수 있으며, recurrent하면 무한 번 계속 반복하게 된다.

  state j 가 i에서 접근가능하고, i가 recurrent하다면 j도 recurrent하다.

- Period of State

  어떤 state i의 period는 다음과 같이 정의된다.

  $$max_d P_{ii}^{n} \neq 0 \leftrightarrow n \equiv_d 0 \tag{53}$$

  이 때, 어떤 state의 period가 1이면 그 state를 aperiodic하다고 한다. recurrent하면서 aperiodic 한 state는 ergodic 하다고 한다.

Limiting Probability 어떤 적절한 행렬 P에 대해서, P_{ii}^{n}가 특정 값으로 수렴할 수 있다. 만약 P가 transition matrix라면, 이를 특정 stochastic process에서 어떤 state로 수렴할 확률로 생각할 수 있을 것이다. 즉,

$$\pi_j = lim_{n\to\infty} P_j^n j \tag{54}$$

를 생각할 수 있다. 이 때, 이 π_j가 수렴할 조건은 이 state j가 ergodic할 때이다. 이 때 π_j들은 다음 연립방정식의 *유일한* 해이다.

$$\begin{aligned} \pi_j &= \sum \pi_i P_{ij} && (55)\\ \sum_j \pi_j &= 1 && (56) \end{aligned}$$

즉, π_j는 **충분한 시간이 지난 후, process가 j번째 state에 있을 확률**이다.

III. Generalization to Random Field

Random Field는 Random Process의 일반화이다. Random Process의 index는 하나의 set이였다면,

III.1 Random Field

III.2 Markov Random Field